DOCUMENT RESUME

ED 278 703 TM 870 129

AUTHOR Moreland, Kevin L.; And Others
TITLE Test Purchaser Qualifications: Present Practice, Professional Needs, and a Proposed System. Issues in Scientific Psychology.
INSTITUTION American Psychological Association, Washington, DC. Scientific Affairs Office.
PUB DATE Jul 86
NOTE 35p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Testing; Ethics; *Evaluation Methods; Models; Professional Associations; Psychological Testing; *Qualifications; Questionnaires; Standards; Testing Problems; Testing Programs; *Test Selection; *Test Use; Test Validity
IDENTIFIERS *American Psychological Association; *Test Publishers

ABSTRACT

Three papers are included in this document. The first, "An Introduction to the Problem of Test User Qualifications," by Kevin L. Moreland, is a short history of the American Psychological Association (APA)/American Educational Research Association (AERA)/National Council on Measurement in Education (NCME) Working Group on Test User Qualifications. The mission and objectives of the group, and the political and practical issues surrounding it are discussed. The second paper, "Establishing Test Purchaser Qualifications: Present Practices and Future Needs," by Gary J. Robertson, reviews the catalog and order fulfillment procedures of 13 major U.S. test publishers in order to determine the types of information required to establish individual purchaser qualifications. It advocates the establishment of a new system with four specific characteristics for evaluating purchaser qualifications. The third paper, "Test Purchaser Qualifications: A Proposed Voluntary System Based on Test Ethics," by Lorraine D. Eyde and Ernest S. Primoff, represents the work of the Test User Qualification Working Group of the Joint Committee on Testing Practices. It also describes a history of the APA's concern with test ethics and test user qualifications, the research conducted by the Working Group, and the preliminary elements of a model voluntary test purchaser system which may be adapted by test publishers. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED278703

TEST PURCHASER QUALIFICATIONS: PRESENT PRACTICE, PROFESSIONAL NEEDS, AND A PROPOSED SYSTEM

ISSUES IN SCIENTIFIC PSYCHOLOGY

Scientific Affairs Office
American Psychological Association

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Boltas

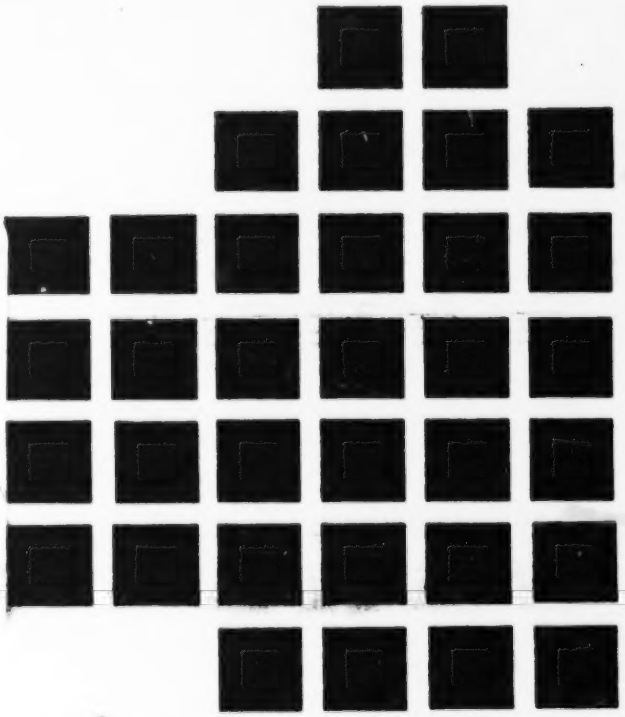
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.



BEST COPY AVAILABLE

TM 880129

TEST PURCHASER QUALIFICATIONS: PRESENT PRACTICE, PROFESSIONAL NEEDS, AND A PROPOSED SYSTEM

An Introduction to the Problem of Test User Qualifications

July 1986

The following paper was written by Kevin L. Moreland of the Professional Assessment Services Division of National Computer Systems. It was adapted from an address presented at the annual meeting of the APA in Washington, D.C., August 26, 1986. Topics covered in the paper are the history of the APA/AERA/NCME Working Group on Test User Qualifications, the mission and objectives of that Working Group, and the political and practical issues surrounding the endeavors of the Working Group. The paper is designed to serve as an introduction to the two papers that follow by Robertson, and Eyde and Primoff.

Inquiries regarding the proposed test purchaser system should be directed to:

Scientific Affairs Office
American Psychological Association
1200 17th Street, N.W.
Washington, D.C. 20036
(202) 955-7755

An Introduction to the Problem of Test User Qualifications

Kevin L. Moreland

NCS Professional Assessment Services
Minneapolis, MN

The Test User Qualifications Working Group traces its origins to the American Psychological Association (APA) Conference on Testing, Assessment, and Public Policy held at the 1981 APA Annual Convention in Los Angeles. One outcome of that conference was a meeting between representatives of the APA and test publishers held in Washington, D.C. in 1984. That meeting covered a broad array of topics of mutual interest to psychologists and test publishers. The two signal concerns identified at that meeting were fair testing practices in education and test user qualifications. A Working Group was formed to deal with each. Both Working Groups now function under the aegis of the APA and two kindred organizations, the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME). Both Working Groups include members of all three organizations. (It is important to note that many of us have a variety of other professional affiliations as well.) The Working Group on fair testing practices presented some results of their deliberations at the AERA meetings held last April, and the activities of the Test User Qualifications Working Group were described during a symposium at the last APA convention. The purpose of this symposium is to discuss the work, to date, of the Test User Qualifications Working Group. In particular, we hope to deal with four questions:

1. What is the problem?
2. What are the issues?
3. What do we know?
4. In what direction should we move?

Mission and Objectives

"Test user qualifications" is a somewhat ambiguous phrase that may take on different meanings for different people. A formal statement of our group's mission and major objectives may therefore help to put the remainder of the symposium in its proper context. At our initial meeting in early 1985 we defined our mission as:

> Consistent with professional standards and ethical principles, to establish a procedure for classifying tests that will lead to a uniform system for assessing user qualifications.

Although we worked very hard on this statement, it may not be clear at this point what we intend our final work product to be. In particular, we don't intend to stop at a "procedure . . . that will *lead to* a uniform system for assessing user qualifications." We intend to go on and develop a model uniform system for assessing user qualifications. Indeed, we are in the midst of that work now, as you will learn momentarily from Dr. Eyde. It is our hope that our model system will be adopted, in whole or part, for the immediate purpose of assessing the qualifications of would-be test users and for the long range purpose of improving the education of would-be test users.

After all this I may still have the cart before the horse; you may still be wondering what the problem is and why we are working on it. I think I can provide some answers without stealing too much of Dr. Robertson's thunder. Consider these specific objectives we set for ourselves:

1. To facilitate access by qualified users to testing materials
2. To facilitate test publishers' procedures for screening

At present, qualification policies are as plentiful as test publishers. A uniform qualification system would spare test purchasers from having to document their qualifications repeatedly. A set of clear guidelines applied to a uniform set of data also would make it easier and more economical for test publishers to do business. Although we haven't set out purposely to build such a system, the ideal test user qualification system should yield a standard form that can be processed by computer.

3. To protect test takers from those who would administer and/or interpret tests they are not qualified to use

The desire to protect test takers is, of course, not new or unique to our group. Some have wanted to protect test takers by banning tests altogether (cf. Amrine, 1965). Government has, at both the state and national levels, enacted laws mandating proper use of tests in education and industry (e.g., Bersoff, 1981). We are simply approaching the same issue from another perspective.

Political Issues

Our efforts to achieve our mission have confronted us with two political issues that bear discussion.

First, as should be obvious to everyone, "test users" are not a homogeneous group. If they were, our group would be out of business because testing would be monitored in much the same way that organized medicine and medical practice laws monitor the practice of surgery. On the contrary, test users can be found in virtually every walk of life. The owner of the clothing store where I buy my suits tests all prospective employees. The upshot of this is that we must be ecumenical in our approach to test user qualifications if we hope to have a broad impact on test use. We realized early on that a parochial approach to the problem would yield, at best, parochial results. Dr. Eyde will elaborate on the steps we have taken to insure broad applicability of our model system for test user qualifications.

In a related matter, we have had to confront issues of civil liberties and civil rights. Any extra-legal attempt to regulate tests in any way could lead to charges that the public's right to freedom of information is being infringed. Would-be test users who are denied access to test materials might also complain that they are the victims of the kind of unfair discrimination that our civil rights laws were enacted to prevent. We have put those issues to rest, as far as our group is concerned, by making it clear that we consider our system to be advisory in nature. Publishers can adopt it or not, in whole or part, as they see fit. On the other hand, we are making no secret of the fact that we hope our model system will enjoy universal application.

Practical Issues

Once again, I may have gotten the cart before the horse. If one does not deal with practical issues in a fair, effective manner, no amount of political ledgerdemain is likely to get one's program adopted.

One obvious practical question is "*Why* are tests misused?" If one hopes to root out misuse it is necessary to know its root causes. Two hypotheses present themselves. Ed Gordon of our Working Group and John Fremer of the Working Group on fair testing practices have hypothesized that one can divide test misusers into two broad groups: venal people who are intentionally up to no good, and well-meaning but misguided people. The collective experience of those of us on the Working Group and of other knowledgeable people whom we've asked indicates that ignorance is the overwhelming cause of test

misuse. This circumstance is fortunate because it allows us to assume the mantle of educators rather than police. We greatly prefer the former role. Besides, we would fail at the latter because we have no enforcement power. This realization has added another objective to those listed above:

4. To improve the education of those who use tests

This objective has two facets. We want to better inform both publishers and users of tests about problems in test use and we want to offer guidance to publishers in the creation of qualification policies for test purchasers. We have come to believe that — in the long run — our educational objective is the most important of the four.

The fact that most test misuse is unintentional also helps us with another problem, namely: How accurate are assertions of qualifications? If we can feel comfortable that people respond as accurately as human foibles permit to questions about their qualifications, this question becomes: How valid is our model system for test user qualifications? The word "valid" is used here in the sense we're all familiar with. In concrete terms, our model system will comprise an algorithm for assigning tests and test services (e.g., computer-based test interpretations) to categories requiring certain test user qualifications, a biodata-type questionnaire for establishing qualifications, and a scoring scheme for the questionnaire. It should be apparent that those components are analogous to the components of an employee selection system and, as such, should be subjected to the same type of scrutiny. If this analogy is not yet apparent to you, it will be after Dr. Eyde's paper.

A final issue that must be dealt with is secondary test distribution. This also has two facets. One is intra-organizational test distribution. All of us have heard horror stories about the psychologist permitting untrained teachers' aides to administer individual intelligence tests — or something equally dismaying. Our model system will not alter such practices in a direct way. Test publishers can control the purchase, but not the subsequent use, of their wares. However, we believe that this practice, too, is largely misguided but not nefarious. This gives us reason to hope that the improvements in the education of test users that may result from our model system will minimize misuse arising from inappropriate distribution of tests within organizations.

Finally, we hope that our work will help to minimize problems arising from secondary test distribution by commercial firms. It appears that the careful attention given to user qualifications by many test publishers goes for nought when they license others to sell their wares. We hope that our work will sensitize publishers to the need to be more careful in this regard. If, as we hope, our work gives rise to relatively uniform qualification policies among publishers, this problem should be minimized even if secondary distribution remains largely a matter of business consideration.

Conclusion

I have dealt briefly with the four questions proposed at the beginning of this paper to give you a flavor of the work of the Test User Qualifications Working Group. Dr. Robertson, Dr. Eyde, and Dr. Primoff will now put some flesh on the skeleton I have given you.

References

Amrine, M. (1965). The 1965 Congressional inquiry into testing: A commentary. *American Psychologist*, *20*, 859-870.

Bersoff, D.N. (1981). Testing and the law. *American Psychologist*, *36*, 1047-1056.

Establishing Test Purchaser Qualifications: Present Practices and Future Needs

July, 1986

This paper was written by Gary J. Robertson, Director of the Test Division, American Guidance Service. The paper discusses current practices of test publishers in screening individual purchasers of educational and psychological tests. Present practices are placed in historical perspective and also are contrasted with control practices followed in such fields as medicine, law, and pharmacy. To ascertain information about present practices, the catalog and order fulfillment procedure of 13 major U.S. test publishers were surveyed in order to determine the types of information required to establish individual purchaser qualifications. The paper advocates the establishment of a new, improved system for evaluating test purchaser qualifications and concludes by delineating four key characteristics desired in such a system.

Establishing Test Purchaser Qualifications: Present Practices and Future Needs

Gary J. Robertson, Ph.D., Director, Test Division
American Guidance Service

There is, at present, an urgent need to re-examine and perhaps to redefine criteria for the sale and distribution of educational and psychological tests. The need for such action became even more apparent following publication of the 1985 *Standards for Educational and Psychological Testing* (APA). Although the *Standards* have been expanded to include new sections devoted to test use, there is little attention directed toward principles to be applied by test publishers in determining appropriateness of purchaser qualifications for a particular type of test. Similarly, the APA "Ethical Principles of Psychologists" (1981) defines, as expected, the role of the psychologist in the use of assessment techniques but offers little assistance in helping to establish qualifications for broader classes of test users. The position taken by this writer and, indeed, by this symposium is that the misuse of tests can be significantly reduced by the development and application of systematic procedures for ascertaining the qualifications of test purchasers for various types of tests; moreover, it is our contention that the implementation of such procedures should be undertaken by test publishers prior to the sale of the desired testing material.

It must be clearly understood that we find the increased responsibility for appropriate test use accorded the user in the 1985 *Standards* both welcome and consistent with the user's role. Standards 5.2, 5.4, and 5.10 (pages 36-37), in particular, seem certain to clarify insofar as the test manual is concerned the uses, limitations, and qualifications needed to administer and interpret a particular test. Standards addressing the "match" between test and purchaser (or user) characteristics are conspicuously absent from the *Standards* and must be addressed, particularly in situations where tests are sold to individuals. As discussed below, the problem of controlling the use of educational and psychological tests is unique when compared with controls used by other professional groups to determine who has access to the "tools" required in a particular profession.

Control Procedures Used in Other Professions

An examination of controls used in other professions such as medicine, law, pharmacy, and dentistry, for example, was conducted in a search for possible models having relevance to the problem of determining who can appropriately use educational and psychological tests. Some interesting contrasts soon became apparent.

In medicine, law, and dentistry controls are really focused primarily on individual practitioners. Licensing of individuals in these professions is a responsibility of the state government. Except for drugs, whose sale is regulated nationally by the Food and Drug Administration, the sale of equipment used by these professionals is uncontrolled. Persons without the requisite licensure may purchase medical or dental equipment and legal books and materials; some medical equipment firms, however, have policies which prohibit the sale of equipment to individuals, but this occurs strictly for business reasons. Pharmacists, however, have legal controls placed on both the material used — drugs — as well as licensure requirements for admittance to the profession. Public pressures undoubtedly were responsible for the enactment of laws governing who can practice in the health-related professions in an attempt to protect the general public from charlatans capable of inflicting serious injury or bodily harm.

Educational and psychological tests, as professional tools, are used by a diverse group of consumers ranging from clinical psychologists to classroom teachers. Such diversity of use precludes a single licensure requirement. In addition, the test materials

themselves vary greatly in complexity and harmful consequences of misuse. For these reasons, the only controls feasible are those placed on the distribution of testing materials at the time of their sale by test publishers.

Present Practices in Screening Purchaser Qualifications

Current practices used by test publishers to govern the sale and distribution of their standardized test instruments can be traced to earlier efforts by the APA Committee on Ethical Standards for Psychology. In 1950, the Committee released its "Ethical Standards for the Distribution of Psychological Tests," which delineated 18 principles governing ethical standards for psychological testing. It was this document that defined a three-level classification system for tests and specified the conditions of governing the release of tests to purchasers. Further impetus toward adoption of the three-level test classification scheme was provided by the 1966 APA publication *Standards for Educational and Psychological Tests and Manuals* prepared by a joint committee representing APA, AERA, and NCME. Standard B.3 stipulated that it was essential to state the qualifications required to administer and interpret a test, and the standard suggested use of the three-level classification system. The two subsequent editions of the *Standards* ceased recommending the three-level classification of tests, but it is clear that the system was found helpful and useful by some publishers as evidenced by the survey described below.

In the absence of further development of specific schemes or recommendations for matching test and purchaser (user) qualifications, this writer investigated a group of test publishers in order to ascertain insofar as possible the degree of control exercised over the release of their test publications. Results of this investigation are described below.

In order to gauge present purchaser screening procedures, a survey was conducted among 13 major test publishers whose testing material accounts for the bulk of test sales to schools, businesses, clinics, private practitioners, and a variety of public and private institutions.* Relevant sections of these publishers' catalogs describing terms for release of materials and, if required, the questionnaire used to assess purchaser qualifications, were examined in order to answer these questions:

1. To what extent do the major test publishers have a general policy which addresses professional qualifications of test purchasers as a condition of sale?
2. What is the professional basis for such qualifications?
3. To what extent do the major test publishers use a detailed questionnaire to determine test purchaser qualifications?
4. What types of information are required on test user qualification questionnaires?

Answers to questions 1,2,and 3 are provided, in part, by Table 1.

* American Guidance Service, CTB/McGraw-Hill, Consulting Psychologists Press, DLM/Teaching Resources, Edits (Educational & Industrial Testing Service), Jastak Assessment Systems, Merrill Publishing Company, National Computer Systems Professional Assessment Services, The Psychological Corporation, The Riverside Publishing Company, Scholastic Testing Service, Science Research Associates, and Western Psychological Services.

Table 1

Type of User Qualification Documentation
Required by Major Test Publishers (N = 13)

Type of Documentation	*N*	%
General Policy of Purchaser Qualifications	13	100.0
Some Documentation of Purchaser Qualifications Required	13	100.0
Detailed Purchaser Qualification Blank Required	7	53.8

Note: Information used to compile this table was obtained from the current publishers' catalogs.

All of the 13 major test publishers stated in their catalogs that they imposed some restriction on the sale of their materials to individuals based on their professional qualifications. Sales to schools and other institutions are made without question if orders are submitted on official stationery or purchase order forms. Conditions of sale often were stated so generally that it was not possible to ascertain the exact criteria used to determine professional eligibility. Considerable variability in the type of information required to establish professional credibility was evident. Twelve of the 13* publishers mentioned adhering to "APA standards," although the exact referent was not clear and it was evident that different publishers interpreted "APA standards" quite differently. For example, one publisher bases its determination solely on professional association membership (i.e., APA, AERA, etc.), several others are nonspecific and simply state that purchasers must establish their professional credentials, and others require detailed information about education and experience. It seems quite safe to infer wide variability in criteria used to establish professional qualifications based on the wide variability in the type of information solicited in the general policy statement.

Seven of the 13 publishers require individuals to complete a detailed questionnaire eliciting specific details about various demographic information used to arrive at a decision about the individual's professional competence. It must be emphasized that this is *inferred* competence based on a variety of background data and not *directly observed* competence.

The types of information asked on the test purchaser qualification questionnaires are shown in Table 2.

* Actually, only 11 specifically mentioned APA standards; the twelfth referred only to "professional organizations."

Table 2

Type of Information Asked on User Qualification Questionnaire by Major Test Publishers (N = 7)

	N	%
Item of Information		
1. Use to be made of tests	5	71.4
2. Educational background/courses	6	85.7
3. Professional association memberships	5	71.4
4. Type of organization where employed	6	85.7
5. Experience	2	28.6
6. References	2	28.6
7. Licensure status	4	57.1

Among the seven publishers who require individuals to complete a special questionnaire to establish their professional credentials, there is evidence of considerable diversity in criteria used to assess purchaser qualifications. As shown in Table 2, there is nearly universal agreement among these firms in requiring information about educational background and the type of organization within which the individual requesting approval is employed. Despite the fact that the specific algorithms used by these seven publishers to evaluate data supplied on the qualification questionnaires are not known, it seems safe to assume here as with the general policy statements that there is considerable variation among test publishers not only with respect to criteria used to determine professional acceptability but also with respect to the weight given to the various data obtained for each potential purchaser.

Future Directions in Assessing Test Purchaser Qualifications

Despite evidence that test publishers do mention adherence to professional standards in their general policies for the sale and distribution of their testing materials, questions remain about the real extent to which various individual purchaser qualifications are assessed by the various publishers and, more specifically, the extent to which there is any degree of consistency across publishers in the criteria used to determine qualifications of test purchasers. The entrepreneurial context within which test publishers exist precludes collaboration or the sharing of information even in an area such as this which concerns professional ethics. Thus, adoption of any type of purchaser qualifications assessment system must be on a purely voluntary basis because there is no effective manner in which to "legislate" a uniform evaluation system. There is, however, a basis for optimism which stems from the fact that all of the major test publishers seemingly evidence concern for the proper use of tests by appropriately trained personnel. Despite rather vague references to "APA standards" by all but one of the test publishing firms, there was information available to indicate that among the very largest publishers, particularly those who publish clinical

instruments, serious attention is paid to screening individual purchasers on a number of relevant criteria.

An examination of the content of the seven test purchaser qualification questionnaires revealed, as stated earlier, both a degree of uniformity as well as diversity. Common sense dictates, to be sure, soliciting information about an individual's professional academic training and about the nature of the work performed, with an emphasis on the role of testing and the use made of test results. A fundamental question remains, however, about the extent to which real professional competence can be judged from an accumulation of courses or an open-ended description of the type of work environment. It would seem that at present purchaser fitness is determined on a very subjective basis by viewing some very concrete, apparent data (e.g., nature and extent of course completed, professional association membership, or licensure status) and some very subjective data (e.g., experience, type of work performed, and occasionally by professional references). There is clearly a need for a system that identifies the most important elements of proper test use and provides a reliable method for ascertaining these key elements. Such a system has been developed and will be proposed in the third paper.

Before consideration of the proposed system for evaluating test purchaser qualifications, there are four desirable characteristics which any sound system for evaluating test purchaser qualifications must possess. Each of these requirements is outlined briefly below. For maximum effectiveness, the system must:

(1) *isolate key elements parsimoniously*. For practical reasons, test purchaser qualification blanks cannot be lengthy and cumbersome. Key elements of critical behaviors important for proper test use must be identified quickly in an efficient manner by as few queries as possible.

(2) *use objective items and decision rules*. To ensure objectivity of outcomes, the items on the questionnaire or other medium used to gather data must lend themselves to objective responses which, in turn, can be evaluated by rapid, objective criteria or decision rules. Such a system can be implemented by persons less highly skilled than professional psychologists, which enhances the likelihood of system acceptance within the entrepreneurial context.

(3) *function well irrespective of order volume*. To be practically feasible, the system must function well in business environments having both large and small order fulfillment characteristics, as well as in situations falling between these two extremes. A system whose complexity precludes its use in large-volume order fulfillment contexts will quickly fall into disuse or will be improperly implemented.

(4) *provide a better match between test and purchaser characteristics*. The present method of ascertaining purchaser characteristics relies on asking general, broad-gauge questions from which enough information can be gleaned to make a determination of fitness for purchase. A system that links critical elements of test use to specific types of tests is highly desirable.

The third paper will propose a screening system which incorporates these desirable features and thus offers a way for test publishers to advance the state of the art of assessing test purchaser qualifications. Furthermore, it is hoped that the proposed system can be used not only by publishers to screen individual purchasers but also by institutional purchasers to help ensure that tests are being used by appropriate personnel in accordance with accepted principles of test use.

References

American Psychological Association. (1950). Ethical standards for the distribution of psychological tests and diagnostic aids. *American Psychologist, 5*, 620-626.

American Psychological Association. (1966). *Standards for educational and psychological tests and manuals*. Washington, D.C.: Author.

American Psychological Association. (1981). Ethical principles for psychologists. *American Psychologist, 36*, 633-638.

American Psychological Association. (1985). *Standards for educational and psychological testing*. Washington, D.C.: Author.

TEST PURCHASER QUALIFICATIONS: A PROPOSED VOLUNTARY SYSTEM BASED ON TEST ETHICS

July 1986

This paper was written by Lorraine D. Eyde, the U.S. Office of Personnel Management, and Ernest S. Primoff, the U.S. Office of Personnel Management (Retired). It is based in part on a paper prepared for presentation at the 1986 Convention of the American Psychological Association, for the symposium, chaired by Robert B. Most, on "Test Purchaser Qualifications: Present Practice, Professional Needs and a Proposed System," Washington, DC, August 26, 1986. The opinions expressed are those of the authors and are not necessarily official policy statements of the U.S. Office of Personnel Management. The paper represents the work of the Test User Qualifications Working Group (TUQWoG) of the Joint Committee on Testing Practices.

The paper provides a history of the American Psychological Association's concern with test ethics and test user qualifications and describes the research conducted by TUQWoG. It describes the preliminary elements of a model, voluntary test purchaser system which may be adapted by test publishers and which has educational implications.

TEST PURCHASER QUALIFICATIONS: A PROPOSED VOLUNTARY SYSTEM BASED ON TEST ETHICS

Lorraine D. Eyde
Ernest S. Primoff, Retired
U.S. Office of Personnel Management

When the public perceives that psychological tests are misused and found to be of questionable reliability and validity, legal restrictions are apt to be imposed on their use. When the polygraph (the galvanic skin response) was limited to its proper use, which was buttressed by a hundred years of research (Woodworth & Schlosberg, 1954), it was considered to be the most widely accepted measure of an active, energized, or tense level of activity. But as the GSR began to be used as a lie detector, critics ignored the advantages of the device, apparent when it was properly used, and moved to limit or ban its use. Questions have been raised not only about the validity of the measure, but also about the adequacy of the psychological training and experience in the interpretation of the psychophysiological measures by those giving polygraph tests (American Psychological Association, 1986, February 1; "Psychologists' Association Says Polygraphs are Unreliable," 1986). As a result, the House voted to outlaw the use of polygraphs in private employment ("A Vote to Ban Lie detectors,"1986) and a bill is under consideration in the Senate ("Lie Detector Lies," 1985), and 26 states and the District of Columbia have statutes restricting the use of the polygraph in the workplace (Bright & Hollon, 1985). If even an extensively researched device like the galvanic skin response can be banned because of improper use, then restrictions on other measurement devices can be expected to follow. For this reason, professional associations concerned with testing and test publishers as well as organizations and individuals using tests need to commit more resources to preventing improper test use. Furthermore, it is recognized that the questions raised about test misuse have major implications for graduate education in the use of tests.

The development of restrictive statutes should alert test publishers and test users to professional and public policy issues relating to the use of a variety of test products currently in wide use. These test products include (a) commercially available tests (see, for example, the tests in Buros' *Mental Measurements Yearbooks*), (b) custom-designed tests, such as teacher certification tests, and (c) computer-based test interpretations (CBTI). Relatively new test products such as CBTI particularly highlight the need for greater regulation of test use (Eyde & Kowal, 1984). For example, computer-based medical expert systems, based on methodologies similar to those used introducing CBTI of psychological tests, have been scrutinized by the Congressional Research Service (Kornbluh, 1985) which recognizes that a number of public policy issues will inevitably warrant congressional consideration. Among these are certification issues and questions about the safety, validity, and reliability of medical expert systems. Possible users of these systems include physicians, paramedical personnel, office personnel, and patients. Among the questions raised are "What kinds of training should [such people] . . . receive to become qualified to operate expert systems? . . . . Should certification standards be established? . . . . Should the Food and Drug Administration (FDA) control such evaluations" (pp. 20-21)? If these medical devices come under the surveillance of the FDA, then it is only a matter of time before computer-based interpretations of psychological tests will also be scrutinized by the government. Thus it becomes critical to develop an improved system for evaluating the competencies of potential CBTI users, before the government regulators raise questions.

This paper summarizes the work of the Test User Qualifications Working Group, one of two working groups of the Joint Committee on Testing Practices (JCTP), which is engaging in such an effort. The JCTP is a cooperative effort with test publishers, which is sponsored by the American Psychological Association (APA), the American Educational

Research Association (AERA), and the National Council on Measurement in Education (NCME). Its objective is to advance, in the public interest, the quality of testing practices.

The purpose of this paper is: (a) to examine the matter of test purchaser and test user qualifications; (b) to document incidents of test misuse; (c) to propose a model system aimed at minimizing test misuse; and (d) to describe preliminary results of our data collection. This proposed voluntary test user qualification system, which may be adapted by different test publishers, is designed to avoid perceptions of restraint of trade. It builds on ethical principles of test use. The paper will describe the process our Working Group has gone through, the steps we have taken, and the steps we anticipate taking in the future, so that interested groups and individuals may become familiar with the project and provide comments to the Working Group.

Current Professional Guidance Dealing with Test Purchaser Qualifications

APA's *Ethical Principles of Psychologists* (APA, 1981) have evolved from critical incidents concerning ethical behavior, originally collected by Nicholas Hobbs and his APA Committee (A. Anastasi letter to L.D. Eyde, July 5, 1985). Golann (1970) reports how the ethics data were collected:

> In 1948, the members of the American Psychological Association, then 7,500 in number, were asked by letter to describe from first hand knowledge a situation in which a psychologist made a decision having ethical implications . . . More than 1,000 reports were received, and edited . . . Members of the committee then drafted statements of standards in the several areas based on the incidents submitted.

As the sections of standards dealing with ethics were finished, they were published in the *American Psychologist*. The first topic to be covered was the sale and distribution of psychological tests and diagnostic aids (APA Committee on Ethical Standards for Psychology, 1950; Hobbs, 1951). The initial code of ethics, developed over a four-year period, was based on principles adopted in 1953 (Golann,1970).

Two of APA's 10 current ethical principles are closely related to the matter of test user qualifications of psychologists and of others who "do work of a psychological nature under the supervision of a psychologist." Principle 2 deals with competence and states that psychologists "only provide services and only use techniques for which they are qualified by training and experience." Principle 8 covers assessment techniques, and states that "Psychologists strive to ensure that the results of assessments and their interpretations are not misused by others."

Furthermore, since 1950 APA has also addressed this subject in test standards (APA, 1954; APA, 1966; APA, 1974; AERA, APA, & NCME, 1985). However, since the issuance of the 1966 standards, the guidance has become more general. Current test standards require that "statements of user qualifications should identify the specific training, certification, or experience needed" to administer a test and interpret it properly (AERA, et al., 1985, p.36). The test standards note that "Responsibility for test use should be assumed by or delegated only to those individuals who have the training and experience necessary to handle this responsibility in a professional and technically adequate manner" (AERA, et al., 1985, p.42).

In addition, the recommendations of the 1981 APA Conference on Testing, Assessment and Public Policy called for developing a method of measuring test user competence (APA, 1981). Still further recommendations grew out of the 1984 APA test publishers' meeting (APA, 1984). As a result of this first major meeting between APA and 23 test publishers and computer interpretation companies, the newly organized Joint

Committee on Testing Practices has responded by creating the Test User Qualifications Working Group.

Development of a Plan

Initial Planning

The work of the Test User Qualifications Working Group began in February 1985,* and focused on a common concern of APA, AERA, and NCME, and the test publishers represented at the 1984 Test Publishers' Meeting. This concern is to develop a system to acquaint prospective test users with the standards they should meet. Such a system should be scientifically, legally, and politically sound, and should meet the business needs of the diverse test publishing community. It should be a model system that test publishers might voluntarily use and adapt to the test products each company sells and distributes. It should not violate antitrust laws affecting the testing industry. In addition to facilitating test publisher procedures for screening test purchasers, the system should facilitate the access of test users to test materials and protect test takers from unqualified professionals who administer and/or interpret tests. Such a system would be built upon ethical principles of testing (see Table 1 for steps in proposed plan).

These steps entailed (1) developing a mission statement, (2) selecting a classification system to cover tests, (3) developing a job analysis-based system with content validity evidence in support of a test user qualification questionnaire, and (4) the collection of job analysis data. The working group is presently working at the third and fourth steps in this process and this paper reflects our effort thus far. The key aspects of our work to date are described in Tables 1 to 4, which (a) summarize the proposed steps to be taken, (b) outline the system for classifying tests with examples of specific tests, (c) give examples of the aspects of test user competence that may be distilled from the job analysis data, and (d) offer an illustration of the questionnaire item types that we might use. This information is given in the sections of the paper describing the initial proposal, data gathering, and further steps.

* The Working Group is made up of the following persons: Lorraine D. Eyde, Chair; Valeria Ford; Edmund W. Gordon; Kevin L. Moreland; Robert B. Most.; and Gary J. Robertson.

Table 1

Steps In Developing A Model Test Purchaser Qualifications System

1. Establish mission statement, emphasizing the relation to ethical principles of testing.
2. Develop a system for classifying tests and selecting examples of tests.
3. Provide a method for developing a system to determine important test user qualifications and to identify evidences of competency, seeking input from test publishers and professional groups at key steps.
4. Develop, collect, and abstract critical incidents and job element information to determine the dimensions of test user competency.
5. Obtain legal consultation.
6. Hold face-to-face meetings with test experts to rate the importance of dimensions of test user competency.
7. Develop a generic questionnaire and possibly questionnaires dealing with specific tests or types of tests, and creating plans (scoring systems) for evaluating prospective users of tests.
8. Prepare report.
9. Disseminate information.
10. Train test publishers in the use of the model voluntary system and facilitate its flexible use.

Table 2

Classification System for Tests with Selected Test Examples

I. Personality

A. Objective Inventories

Minnesota Multiphasic Personality Inventory (University of Minnesota Press); California Psychology Inventory (CPP); 16 Personality Factor Questionnaire (IPAT); Myers-Briggs Type Indicator (CPP); State-Trait Anxiety Inventory (CPP).

B. Interests, Values, and Personal Orientations

Strong-Campbell Interest Inventory (Stanford University Press); Kuder Occupational Interest Survey (SRA); Bem Sex-Role Inventory (CPP); Self-Directed Search (PAR); Activity Vector Analysis (Walter V. Clarke Associates).

C. Projective Tests

Rorschach (Hans Huber); Thematic Apperception Test (Harvard U); Rotter Incomplete Sentence Test (PC).

D. Other

Reid Report (Reid Psychological Systems).

II. Tests of Separate Abilities: Occupational

A. Proficiency and Skill Tests

SRA Typing Skills (SRA)

B. Aptitude Tests

Armed Services Vocational Aptitude Battery (U.S. Military Enlistment Processing Command); Differential Aptitude Tests (PC); Short Employment Tests (PC); Purdue Pegboard (SRA); General Aptitude Test Battery (USES); Wesman Personnel Classification Test (PC); Bennett Mechanical Comprehension Test (PC); SRA Mechanical Aptitudes (SRA).

III. Educational and Achievement Tests

California Achievement Tests (CTB); Iowa Tests of Basic Skills (Riverside); Iowa Test of Educational Development (SRA); SRA Achievement Series (SRA); Tests of Achievement and Proficiency (Riverside); Wide Range Achievment Test (Jastak); Peabody Individual Achievement Test (AGS); Stanford Diagnostic Mathematics Test (PC); Stanford Diagnostic Reading Test (PC); Nelson Reading Skills Test (PC); Metropolitan Readiness Tests (PC); Boehm Test of Basic Concepts (PC); Slosson Oral Reading Test (Slosson).

IV. Neuropsychological

Halstead-Reitan Neuropsychological Test Battery (Reitan Associates); Luria-Nebraska Neuropsychological Battery (WPS): Bender Visual-Motor Gestalt Test (American Orhthopsychiatric Association); Wechsler Memory Scale (PC); Goldstein-Scheerer Tests of Abstract and Concrete Thinking (PC).

V. Speech, Language, and Hearing

Porch Index of Communicative Ability (CPP), Goldman-Fristoe-Woodcock Auditory Skills Test Battery (AGS); Goldman-Fristoe Test of Articulation (AGS); Illinois Test of Psycholinguistic Abilities (Univ. of IL Press); Khan-Lewis Phonological Test (AGS).

VI. Tests of General Intellectual Level

A. Individual Intelligence Tests

Kaufman Assessment Battery for Children; Slosson Intelligence Test (Slosson); Wechsler Adult Intelligence Scale-Revised (PC); Wechsler Intelligence Scales for Children - Revised (PC); Stanford-Binet Intelligence Scale-Revised (Riverside); Peabody Picture Vocabulary Test (AGS), Full Range Picture Vocabulary Test (Psychological Test Specialists).

B. Group Intelligence Tests

Cognitive Abilities Test (Riverside); Otis-Lennon School Ability Test (PC); Culture Fair Intelligence Test (IPAT); Raven's Progressive Matrices (PC); School and College Ability Test (ETS/CTB); Test of Educational Ability (SRA); Test of Cognitive Skills (ETB/ McGraw-Hill).

VII. Development Scales

Adaptive Behavior Scale for Children and Adults (AAMD); System of Multicultural Pluralistic Assessment (PC); Denver Developmental Screening Test (LADOCA); Cain-Levine Social Competency Scale (CPP); Vineland Adaptive Behavior Scales (AGS); Bayley Scales of Infant Development (PC); Developmental Indicators for the Assessment of Learning (Grolier).

Note: See Appendix A for full names of test publishers.

Table 3

Examples of Possible Elements, their Generic Subelements, and Subelements for Specific Tests : Preliminary Job Analysis Data

---

Element 1. General Test User Competence

a. *General subelements*

Acceptance of responsibility for competent use of the test. Not making copies of copyrighted materials.

b. *Subelements for specific tests*

Knowledge of manual and related user's guides for Test X.

c. *Evidences*

Do you agree not to make copies of copyrighted test materials? Do you have a copy of the manual for the test you are ordering and are you familiar with it?

Support your answers by referring to your education, experience, informal training and experience, etc. given in the test user questionnaire.

Element 2. Avoiding Overemphasis of a Test Result

a. *Generic subelement*

Knowledge of need for multiple sources of convergent data.

b. *Specific subelement*

Using scores on intelligence test to make vital decisions about a person's life only in conjunction with other pertinent affective or interest-related variables, vocational aptitude, neurological functioning, creativity, etc.

c. *Evidences*

Do you plan to consider the test result for Test X as only one evidence to be evaluated before making a decision or counseling someone? Support your willingness and competence to do this by referring to courses, workshops, etc. which you have taken or given, to books and manuals you have studied or written, and to intern and fully professional experience. You may refer to items in your user questionnaire.

Element 3. Interpreting Test Results

a. *Generic subelements*

Understand norms and limitations.
Knowledge of test theory and principles of interpretation.
Consideration of standard error of measurement.
Considering whether local purpose of test use meets the purpose for which the test was intended.

b. *Specific subelements*

AAMD Adaptive Behavior Scales: Recognizing situational effects when teacher's ratings disagree with parents.
Myers-Briggs Type Indicator: Recognition that scores are preference scores, not ability.

Rorschach: Actually scoring the test, rather than getting only an impression of behavior.

c. *Evidences*

Do you understand the theory of Test X? Tell about how well you keep up with the test literature, workshops attended, etc.

Do you compare your interpretations of test scores with those of other professionals?

Element 4. Accuracy

a. *Generic subelements*

Establishing rapport with examinees so as to obtain accurate scores. Checking frequently during scoring to catch lapses.

b. *Specific subelements*

Wechsler Adult Intelligence Scale-Revised: Using prescribed wording, not using "clinical sense" to change wording, giving feedback, or using illegal cues and queries.

c. *Evidence*

Do you keep a check on accuracy of scoring? If so, how?

Element 5. Providing Meaningful Interpretation of Test Results

a. *Generic subelements*

Interpreting test scores to parents and teachers, rather than simply transmitting scores labeling child without consideration of compensating strengths and actual school performance.

b. *Specific subelements*

Wechsler Intelligence Scale for Children-Revised: Refraining from giving actual test scores and computerized summaries directly to lay persons seeking information about test results.

c. *Evidences*

Do you plan to give test score feedback and interpretation? If so, explain how, by whom, using what formats, and what amount of time.

Table 4

Illustration of Possible Questionnaire Items for Use in Linking Education, Training, and Experience to Ethical Test Use

---

Formal Education and On-the-Job Training (includes taking workshops)

Relevant Courses	*Dates*	*Institution*	*Tests Studied*
A			
B			
C			
D			
E			

Experience

(Includes volunteer professional activities, e.g., testing and ethics committees, test reviews, professional papers, and workshops given)

Experience	*Dates*	*Institution*	*Tests Studied*
1			
2			
3			
4			
5			
6			
7			

Scale for assessing knowledge, skill, or ability:

A. I have little or no training or experience in this.
B. I have been trained in this or studied it myself.
C. I have done this in my work or professional activities.
D. I feel I can do this without supervision.

Scale for assessing willingness to engage in activity:

W. Willing to do it.
U. Unwilling to do it.

Early in our discussions, we considered basing our system on the degree of harm to test takers which might be caused if the test were improperly used for a particular function. The possible range of uses included (a) self-edification, (b) screening, (c) early identification, (d) diagnosis, (e) individual evaluation, (f) instructional planning, (g) prognosis, (h) treatment, (i) program evaluation, (j) research, and (k) classification. This idea was shelved because the implementation of the four steps outlined above would require considerable effort. The consideration of the impact of different uses of a test would add an additional multiplier to each dimension and make the proposed system impractical.

Outreach Activities

In order to gain test publisher views, we sent our statement of goals and requirements to a sample list of 198 test publishers and computer-based test interpretation services, developed largely from the draft list of test publishers used by the Buros Institute's *Ninth Mental Measurements Yearbook*. Included in this list were all USA and Canadian companies that published four or more tests, as well as a sample of companies that published one to three tests. We added to this list those computer-based interpretation companies that attended the 1984 APA-sponsored meeting of test publishers.

Our plans also called for working with allied professional groups who use tests in their professional work. Mailing lists of the Joint Committee on Testing Practice, for example, have included many professional organizations and numerous organizations concerned with licensing and certification examinations. Our Working Group realized that test users represent many different occupational groups, and we agreed that our proposed system should not eliminate any potential test purchaser solely on the basis of occupational titles. Instead, we decided that the system should identify the competencies of individual test purchasers, evaluating not only their education and training, but also their ethical test practices.

Response from Test Publishers

Our Working Group was pleased with the response to our request for test publisher input into our planning process. Approximately one-fourth (47) of the companies replied and their responses were used to develop our plan. The following examples highlight comments used in shaping our proposal. One dealt with the impact of computers on test use and the other with the broad range of test users who use the tests after they are purchased.

For example, a statement from Thoughtware, a company that produces com;uter-based training aids, reminded us of the need for covering software products for our system. The company (James R. Caplan letter to D. Boltas, May 6, 1985) stressed that "the use of computers will profoundly change the requirements of those who use psychological tests to help make decisions about individuals." We agree and computer-based test products are included in our comprehensive definition of tests, which covers psychological, educational, and employment tests that have standardized procedures for test administration or scoring or that use normative data for making interpretations. The system is designed so that custom-designed tests, such as teacher certification tests, may be included at a later date.

A second example is a reminder of the important distinction between "test purchaser" and "test user." Test publishers sell tests to individual purchasers. These persons then often delegate test use to others. Scott D. Thompson of the National Association of Secondary School Principals (letter to D. Boltas, May 8, 1985) emphasized concerns about who makes use of tests. He noted that "A user qualification system has been needed for some time now, as the present system simply regulates who may purchase different categories of tests." Our Working Group and the JCTP are well aware of the

concerns raised about the qualifications of the broad range of professionals who use the tests that organizations purchase. While the primary concern of this Working Group is to develop a model test purchaser system, our long-term objective is to address the complex matter of the qualifications of those who use tests after they are purchased.

Documenting Test Misuse

Both the short- and long-term objectives deal with the impact that tests have on test takers. Thus our project is a consumer effort aimed at reducing test misuse. Before beginning to develop our system, therefore, we reviewed existing evidence of test misuse.

Test Misuse Complaints

Complaints about test misuse may be brought before the APA Ethics Committee, State Psychology Boards, and State Psychological Associations., Test misuse issues have come before these groups. For example, in 1983 the APA Ethics Committee had under investigation five cases of test misuse (Mills, 1984), including one related to CBTI (D.H. Mills letter to L.D. Eyde, July 17, 1984). The APA's Committee on Professional Standards (1985) conducted a survey of test misuse, to which 49 State Psychology Boards responded. Of the total of 776 complaints received by the responding Boards, over a specified 12-month time period, 44 (or 5.7 %) included allegations of improper use of tests by psychologists. Of these 44 complaints, 5 (or 11%) had resulted or were expected to result in a hearing (B. Wand letter L.D. Eyde, December 5,1985).

Initial Proposal

Because of our concern over the impact that tests have on test takers, our Working Group decided to base our system on critical incidents (Flannagan, 1954) of test misuse. In addition, the principles of the Job Element Method (Primoff, 1975; Primoff, Clark & Caplan, 1982) are used to assess the minimum requirements for competent test use. These two methods identify the important dimensions of test user qualifications.

We developed a system for classifying tests that relied heavily on the system used by Anastasi (1982) (see Table 2). This list includes specific examples of tests, covering many of the most frequently used tests listed by Lubin, Larsen, Matarazzo (1984), tests listed by Anastasi, and unusual tests. We included tests from companies that publish a large number of tests as well as from those that publish a smaller number of tests. The list also reflects the availability of raters familiar with the use of the test.

Next we developed a working paper outlining how Job Element Method principles might be applied in our proposed system. (Critical incidents would be gathered in accordance with Flannagan's technique, and the Job Element Method, originally designed for employee selection, would require some adaptation.) The Job Element Method has been described by Anastasi (1982), McCormick (1979), and others, and was used by Beier and Ekstrom (1979) to develop a system for evaluating competencies acquired through volunteer experience. The method focuses on what people can do and are willing to do and does not rely solely on credentials. It builds on the knowledge and experience of subject matter experts (a) to identify important requirements, (b) to establish procedures for obtaining information from applicants (such as a self-report form), and (c) to develop a crediting plan for scoring evidence of competency. It concentrates on assessing individual differences in competency.

Once we had developed the working paper and plan for the development of the test purchaser system, we solicited comments from psychologists with expertise in this area. We decided to start by collecting job analysis data from test experts in the use of specific tests, using a rater form. The instructions to raters and the rater form are given in Appendix B.

Data Gathering and Preliminary Results.

Data collection is now in progress. The rater forms are being used to obtain information that may be used to determine the dimensions of test user competency and to identify evidences of test user competency. The rating forms are being sent to persons who have reviewed the tests for the Buros Institute or the Test Corporation of America, to textbook writers, to recognized trainers, or to professionals who are employed by test publishers, and to practitioners.

As these data are obtained, the broad dimensions ("elements" in Job Element Method terminology), and their more specific dimensions ("subelements") are abstracted.

The job analysis data we have collected thus far yields five tentative elements (see Table 3). These five elements are: (1) General Test User Competence, (2) Avoiding Overemphasis of a Test Result, (3) Interpreting Test Results, (4) Accuracy, and (5) Providing Meaningful Interpretation of Test Results.

These broad test user dimensions have, on the bases of job analyses information, been found to apply to a broad range of tests. Preliminary analysis of our job analysis data shows that some requirements such as generic subelement number 3 "Not making copies of copyrighted materials" may apply to tests in general, whereas others such as number 4 "Knowledge of manual and related user's guides for Test X" relate to specific tests.

The element on "Avoiding Overemphasis of a Test Result" is related to general subelement number 8, "Knowledge of need for multiple sources of convergent data." Thus far we have found related subelements, for example, for these instruments: California Psychological Inventory, the Myers-Briggs Type Indicator, the Minnesota Multiphasic Personality Inventory, the Rorschach, the System of Multicultural Pluralistic Assessment (SOMPA), and the Wechsler Intelligence Scale for Children-Revised.

The elements and subelements listed in Table 3 may be used to develop questions for the test user questionnaire to determine whether or not a person is qualified to purchase a test or a group of tests. Table 3 lists some questions that may be used to gather evidence (see questions 5 through 7) that potential test purchasers meet the requirements related to Element 1, "General Test User Competence." These might include: Number 6, "Do you have a copy of the manual for the test you are ordering and are you familiar with it?"

Our critical incidents and the resulting test user dimensions may form the basis for our test user qualifications system. Possible additional questionnaire items to be used by potential test purchasers to permit linking education, training, and experience to ethical test use are included in Table 4. The subelements, evidences, and questionnaire items may also provide university personnel and trainers with job analysis data on which to base educational and training programs concerning the use of tests.

Note that the evidences in Table 3 and the items in Table 4 illustrate the broad range of information that might be elicited from potential test purchasers. Evidences might be obtained from multiple sources, including traditional credentials and nontraditional examples of competency such as educational and other life experiences, work experience, licenses, test scores, or academic and employment references. Included might be (1) courses in measurement, psychodiagnostics, or psychopathology, (2) state licenses, (3) certification by the American Board of Professional Psychologists Association (and other professional associations). The following types of evidences might also be covered: (1) efforts at continuing education, including self-study and tutoring of others, (2) making presentations at professional association conventions, publishing relevant papers and test reviews, (3) conducting workshops and graduate-level courses and seminars, and (4) peer review of psychological test reports. Evidence relating to ethics and malpractice complaints may also be considered.

Our experts have been providing us not only with the requested job analysis data, as in the example in Table 3, but also with comments on educational problems. Here are quotations from three experts:

- Certainly completion of an accredited school or clinical psychology program should be adequate protection against these misuses, yet I see many people who have completed accredited programs . . . guilty of . . . these . . . misuses (A.S. Kaufman on the WAIS-R).

- There is scant TAT [Thematic Apperception Test] training in most APA-approved clinical programs. I do not believe that being a professional psychologist is necessarily related to TAT competence. (R.H. Dana letter to K.L. Moreland, July 25, 1985).

- I feel that the problem we are facing concerning abuse of tests and abuse of computer programs for tests derives from the same core; namely training.

  Beginning in the early 1960's and extending well into the 1970's most training programs, regardless of the specialty, that is, clinical, school, organizational, etc., have focused primarily on intervention tactics and unfortunately, the major emphasis on intervention models has been at the training expense of assessment procedures. During the 1940's and 1950's, assessment procedures as such constituted the major segment of doctoral training in most specialty areas and especially clinical psychology. Four or five courses concerning assessment procedures were not uncommon and the skill levels that were developed among students were clearly manifest in the product. But that is no longer the case. Many universities devote no more than two semesters. Of course this is enhanced by practicum experience, but the typical new graduate tends to flounder in most areas of assessment. (John E. Exner, Jr., letter to K.L. Moreland, August 31,1985 on the Rorschach.)

Further Steps

Our future plans call for seeking legal consultation on the development of the proposed system. Once the collection of rating data is completed and the test user dimensions are abstracted, we expect to convene a panel of test experts to rate the importance of the test user dimensions (elements and subelements). Four criteria, used in the Job Element Method, may be applied: (a) how frequently the elements appear in acceptable test users who are at the marginal level, (b) the extent to which the element identifies superior test users, (c) the extent to which the elements might cause trouble if they are not considered, and (d) the extent to which they are practical to use in assessing competencies. Statistical values may be obtained for these criteria. These data could be made available to test publishers for each to choose the elements most suitable for their specific tests.

Our voluntary proposed system is designed to be flexible and adaptable to the needs of individual test publishers. The questionnaire items that we develop may be generic, that is, cover many types of test use, or they may relate to specific tests. The data we have gathered thus far appear to indicate that we will develop a consciousness-raising questionnaire, to be completed by potential test users, which will identify key test user dimensions based on test ethics and relate these to their training and experience. We also expect to obtain information on ethical test use from endorsers of individual potential test purchasers. In addition, suggestions will be made for verifying information about course completion, degrees received, and job experience. Also we expect to prepare a report and disseminate information.

In summary, this paper describes the need for minimizing test misuse and describes the early stages of the development of a potential test purchaser qualification system. This system would be a data-based system as a model system, parts of which may be adapted to

the needs of individual test publishers. The data on test misuse may also be used to strengthen graduate programs to prepare psychologists and allied professional groups in the ethical use of tests.

References

American Educational Research Association, American Pychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing* . Washington, DC: American Psychological Association.

American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. Supplement. *Psychological Bulletin, 51*, 210-238.

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. (1966). *Standards for educational and psychological tests and manuals* . Washington, DC: American Psychological Association.

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. (1974). *Standards for educational pychological tests* . Washington, DC: American Psychological Association.

American Psychological Association. (1981). *Ethical principles of psychologists* (Rev.ed.). Washington, DC: Author.

American Psychological Association. (1981, December). *Report on the Conference on Testing, Assessment and Public Policy*, August 25, 1981, Washington, DC. Author.

American Psychological Association. (1984, July 13-16). *Report on test publishers' meeting* . Washington, DC: Author.

American Psychological Association Committee on Professional Standards. (1985). *Survey: Reported Misuse of Tests* . Agenda Item No.19, Committee on Psychological Tests and Assessment, April 11-13, 1985.

American Psychological Association. (1986, February 1). APA Resolution Says Reliability of Polygraph Test "Unsatisfactory." Press Release. Public Information Office.

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: MacMillan.

APA Committee on Ethical Standards for Psychology. (1950). Ethical Standards for the distribution of psychological tests and diagnostic aids: Code of Standards for Test Distribution, *American Psychologist*, 5, 620-626.

"A vote to ban lie detectors". (1986, March 24). *Time* , 39.

Beier, J.J., & Ekstrom R.B. (1979). Creating employment equity through recognition of experiential learning. *Journal of Career Education*, 6, 2-11.

Bright, T.L., & Hollon, C.J. (1985, February). State regulation of polygraph tests at the workplace. *Personnel*, 62 (2), 50-56.

Eyde, L.D., & Kowal, D.M. (1984). Ethical and professional concerns regarding computerized test interpretation services and users. Paper from symposium on "The Use of Computer-Based Test Interpretations: Prospects and Problems." Joseph D. Matarazzo, Chair, August 27, 1974, Toronto, Canada.

Flannagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51 (4), 327-358.

Golann, S.E. (1970). Ethical standards for psychology: Development and revision, 1938-1968. *Annals of the New York Academy of Sciences*,169 (2), 398-405.

Hobbs, N. (1951). Report of the American Psychological Association's Committee on Ethical Standards for Psychology to the Council of Representatives. Washington, DC: American Psychological Association.

Kornbluh, M. (1985). Medical practice and artificial intelligence: An odd couple? *Congressional Research Service Review, 6* (10), 19-21.

"Lie detector lies." (1985, August 8). *The New York Times*, 20E.

Lubin, B., Larsen, R. M., & Matarazzo, J.D. (1984). Patterns of psychological test usage in the United States: 1935-1982. *American Psychologist*, 40, 451-454.

McCormick, E.J. (1979). *Job analysis: Methods and applications*. New York: AMACOM.

Mills, D.H. (1984). Ethics education and adjudication within psycholgy. *American Psychologist*, 39, 669-675.

Primoff, E.S. (1975). *How to prepare and conduct job element examinations*. U.S. Civil Service Commission, Personnel Research and Development Center, Technical Study 75-1. Washington, DC: U.S. Government Printing Office, 1975. (NTIS No. PB 84-166 743).

Primoff, E.S., Clark, C.L., & Caplan, J.R. (1982). *How to prepare and conduct job element examinations: Supplement*. U.S. Office of Personnel Management, Offfice of Personnel Research and Development Report 82-3. Washington, DC. (NTIS No. PB 84-148 162).

"Psychologists' Association Says Polygraphs are Unreliable." (1986, February 9). *The New York Times*, 26.

Woodworth, R.S., & Schlosberg, H. (1954). *Experimental psycholgy*. New York: Holt.

Appendix A

List of Abbreviations and Names of Test Publishers

AAMD: American Association of Mental Deficiency
AGS: American Guidance Service
CPP: Consulting Psychologists Press, Inc.
CTB: California Test Bureau/McGraw-HIll
ETS: Educational Testing Service
Grolier: Grolier Educational Corporation
Harvard U: Harvard University Press
IPAT: Institute for Personality and Ability Testing, Inc.
Jastak: Jastak Associates, Inc.
LADOCA: LADOCA Publishing Foundation
PAR: Psychological Assessment Resources, Inc.
PC: The Psychological Corporation
Reitan: Reitan Neuropsychological Laboratories
Riverside: Riverside Publishing Co.
Slosson: Slosson Educational Publications, Inc.
SRA: Science Research Associates, Inc.
USES: U.S. Employment Service

Note: For addresses, see James V. Mitchell, Jr. (Ed.) (1985). *The Ninth Mental Measurements Yearbook* (Vol.II). The University of Nebraska Press, The Buros Institute of Mental Measurements: Lincoln, or Anastasi, A. (1982). *Psychological Testing,* (Fifth Edition). New York: Macmillan.

Appendix B. Cover Memo and Rater Form

Test User Qualifications Working Group (TUQWoG)

Date:
From:
Phone number and address:

To:

Test:
Date of return of the attached form:

The Test User Qualifications Working Group of the Joint Committee on Testing Practices requests your cooperation in developing a systematic procedure to assess the qualifications of test purchasers. This project is consistent with the mission of the Joint Committee to advance, in the public interest, the quality of testing practices. An adaptation of the *Job Element Method of Examining* (JEM) constitutes the methodology being used to formulate the proposed user qualifications for various types of educational and psychological tests. The following activities, based upon the Job Element Method are planned:

Phase I:

Experts will be asked to list examples of misuse for one or two representatives of major types of standardized tests. This phase will generate a listing of critical incidents of misuse for all classes of standardized tests in use at the present time.

Phase II:

Data from Phase I will form the basis for developing a structured rating form which will be used to rate a broad sampling of tests of all types.

Phase III:

Results of Phase II ratings will be used to develop a classification scheme to define essential qualifications needed by purchasers of various types of tests. Results of Phase III will be disseminated to all publishers for their voluntary use in determining user qualifications. It is hoped that the implementation of this system will ultimately result in more uniform standards for determining test user qualifications which will reduce the misuse of tests and thereby benefit all test users and all test takers.

We are writing to you as a recognized expert to assist with Phase I of this project. You are requested to complete the attached form for the test listed above which is being used in Phase I of the research project. Please follow the specific directions carefully. If you have any questions, please do not hesitate to phone me at the above number. We would appreciate receiving this information by the date listed above. Thank you for your cooperation with Phase I of our research. Determining test user qualifications has in the past been a rather arbitrary process with little to guide test publishers. It is hoped that the results of this research will provide a more objective basis which test publishers may elect to implement within their own framework for determining test user qualifications.

(6) We have gathered information from several other expert professionals on examples of test misuse and the knowledges, skills, and abilities that will prevent the misuse(s). For example, education and training has emerged as one category where we need more specific information about the *specific* knowledges that will prevent a particular misuse.

Consider the misuse you gave above and list below *one* specific knowledge, skill, ability, or other characteristic for each broad category of measurement knowledge. *Be as specific as you can for each knowledge category.* Also list the potential harm that could come from the *lack* of the category of measurement knowledge.

Example: Administration (category) —"Ability to establish rapport with the examinee." Potential harm — examinee anxiety about testing process and invalid test results.

Please add any additional content categories in the blanks provided at the end of the list if you feel there are omissions.

(1) Administration _____

Potential harm _____

(2) Scoring _____

Potential harm _____

(3) Norms _____

Potential harm _____

(4) Reliability _____

Potential harm _____

(5) Validity _____

Potential harm _____

(6) Interpretation _____

Potential harm _____

(7) _____

Potential harm _____

3/1/86

Rater Form

Test User Qualifications Working Group (TUQWoG)

Date_____
Form Requested by:_____

Rater Form

Print your name _____
Address _____
_____

Rating of (Name of Test)_____

Please provide the following information for each example of actual test misuse you have observed. Be as specific as possible.

(1) Describe the example of test misuse.*

(2) Indicate what led to this misuse.

(3) What potential harm could this misuse have caused to an individual?

(4) What test user characteristics [knowledges, skills, abilities, and other characteristics (KSAO)] would have reduced the chances of this misuse?

(5) Indicate types of evidences of competency that might be used by test purchasers to document these KSAO's. List ways that test purchasers can document these KSAO's through formal education, job experience, informal life experience, or any other means.

* Please provide us with as many different examples as possible, listing them as example A, example B, etc.